arXiv:0712.3650v2 [math.PR] 14 Nov 2008

Large deviations for eigenvalues of sample covariance matrices, with applications to mobile communication systems

Anne Fey [*] Remco van der Hofstad[†] [†] Marten J. Klok [‡]

November 17, 2008

Abstract

We study sample covariance matrices of the form $W = \frac{1}{n}CC^T$, where C is a $k \times n$ matrix with i.i.d. mean zero entries. This is a generalization of so-called Wishart matrices, where the entries of C are independent and identically distributed standard normal random variables. Such matrices arise in statistics as sample covariance matrices, and the high-dimensional case, when k is large, arises in the analysis of DNA experiments.

We investigate the large deviation properties of the largest and smallest eigenvalues of W when either k is fixed and $n \to \infty$, or $k_n \to \infty$ with $k_n = o(n/\log\log n)$, in the case where the squares of the i.i.d. entries have finite exponential moments. Previous results, proving a.s. limits of the eigenvalues, only require finite fourth moments.

Our most explicit results for k large are for the case where the entries of C are ± 1 with equal probability. We relate the large deviation rate functions of the smallest and largest eigenvalue to the rate functions for independent and identically distributed standard normal entries of C. This case is of particular interest, since it is related to the problem of the decoding of a signal in a code division multiple access system arising in mobile communication systems. In this example, k plays the role of the number of users in the system, and n is the length of the coding sequence of each of the users. Each user transmits at the same time and uses the same frequency, and the codes are used to distinguish the signals of the separate users. The results imply large deviation bounds for the probability of a bit error due to the interference of the various users.

Key words: Sample covariance matrices, large deviations, eigenvalues, CDMA with soft-decision parallel interference cancelation.

1 Introduction

The sample covariance matrix W of a matrix C with k rows and n columns is defined as $\frac{1}{n}CC^T$. If C has random entries, then the spectrum of W is random as well. Typically, W is studied in the case that C has i.i.d. entries, with mean 0 and variance 1. For this kind of C, it is known that when $k, n \to \infty$ such that $k/n = \beta$, where β is a constant, the eigenvalue density tends to a deterministic density [19]. The boundaries of the support of this distribution are $(1-\sqrt{\beta})_+^2$ and $(1+\sqrt{\beta})^2$, where $x_+ = \max\{0, x\}$. This suggests that the smallest eigenvalue $\lambda_{\min}$ converges to $(1-\sqrt{\beta})_+^2$, while the largest eigenvalue $\lambda_{\max}$ converges to $(1+\sqrt{\beta})^2$. Bai and Yin [4] have proved a.s. convergence of $\lambda_{\min}$ to $(1-\sqrt{\beta})_+^2$. Bai, Silverstein and Yin [3] proved a.s. convergence of $\lambda_{\max}$ to $(1+\sqrt{\beta})^2$, see also [23]. The strongest results apply in the case that all entries of C are i.i.d. with mean 0, variance 1 and finite fourth moment. Related results, including a central limit theorem for the linear spectral statistics, can be found in [1, 2], to which we also refer for an overview of the extensive literature.

In the special case that the entries of C have a standard normal distribution, W is called a *Wishart matrix*. Wishart matrices play an important role in multivariate statistics as they describe the correlation structure in i.i.d. Gaussian multivariate data. For Wishart matrices, the large deviation

[*]Delft Institute of Applied Mathematics, Faculty of EEMCS, Delft University of Technology, The Netherlands, a.c.fey-denboer@tudelft.nl

[†]Department of Mathematics, Eindhoven University, of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands, rhofstad@win.tue.nl

[‡]ORTEC BV, Orlyplein 145c, 1043 DV Amsterdam, The Netherlands, MKlok@ortec.nl

rate function for the eigenvalue density with rate $\frac{1}{n^2}$ has been derived by Guionnet [9] and Hiai and Petz [11]. However, the proofs depend heavily on the fact that C has standard normal i.i.d. entries, for which the density of the ordered eigenvalues can be explicitly computed.

In this article, we investigate the large deviation rate functions with rate $\frac{1}{n}$ of the smallest and largest eigenvalue of W, for certain non-Gaussian entries of C. We pose a strong condition on the tails of the entries, by requiring that the exponential moment of the square of the entries is bounded in a a neighborhood of the origin. We shall also comment on this assumption, which we believe to be necessary for our results to apply.

We let $n \to \infty$, and k is either fixed or tends to infinity not faster than $o(n/\log\log n)$. Our results imply that all eigenvalues tend to 1 and that all other values are large deviations. We obtain the asymptotic large deviation rate function of $\lambda_{\min}$ and $\lambda_{\max}$ when $k \to \infty$. In certain special cases, we can show that the asymptotic large deviation rate function is equal to the one for Wishart matrices, which can be interpreted as saying that the spectrum of sample covariance matrices with k and n large is close to the one for i.i.d. standard normal entries. This proves a kind of universality result for the large deviation rate functions.

This paper is organized as follows. In Section 2, we derive an explicit expression for the large deviation rate functions of $\lambda_{\min}$ and $\lambda_{\max}$. In Section 3, we calculate lower bounds for the case that the distribution of C_{mi} is symmetric around 0, and $|C_{mi}| < M$ almost surely, for some $M > 0$. In Section 4, we specialize to the case where $C_{mi} = \pm 1$ with equal probability, which arises in an application in wireless communication. We describe the implications of our results in this application in Section 5. Part of the results for this application have been presented at an electrical engineering conference [7].

2 General mean zero entries of C

In this section, we prove large deviation results for the smallest and largest eigenvalues of sample covariance matrices.

2.1 Large deviations for $\lambda_{\min}$ and $\lambda_{\max}$

Define $W = \frac{1}{n}CC^T$ to be the matrix of sample covariances. We denote by $\mathbb{P}$ the law of C and by $\mathbb{E}$ the corresponding expectation. Throughout the paper, we assume that the i.i.d. real matrix elements of C are normalized, i.e.,

$$\mathbb{E}[C_{ij}] = 0, \qquad \mathrm{Var}(C_{ij}) = 1. \tag{1}$$

The former implies that a.s., the off diagonal elements of the matrix W converge to zero, the second implies that the diagonal elements converge to 1, a.s. By a rescaling argument, the second assumption is without loss of generality.

In this section, we rewrite the probability for a large deviation of the largest and smallest eigenvalues of W, $\lambda_{\max}$ and $\lambda_{\min}$, respectively, into that of a large deviation of a sum of i.i.d. random variables. This rewrite allows us to use Cramér's Theorem to obtain an expression for the rate function. This section gives a heuristic derivation of our result, that will be turned into a proof in Section 2.2.

For any matrix W, and any vector $\mathbf{x}$ with k coordinates and norm $\|\mathbf{x}\|_2 = 1$, we have

$$\lambda_{\min} \le \langle \mathbf{x}, W\mathbf{x}\rangle \le \lambda_{\max}.$$

Moreover, for the normalized eigenvector $\mathbf{x}_{\min}$ corresponding to $\lambda_{\min}$, the lower bound is attained, while for the normalized $\mathbf{x}_{\max}$, corresponding to $\lambda_{\max}$, the upper bound is attained. Therefore, we can write

$$\begin{aligned} P_{\min}(\alpha) &= \mathbb{P}(\lambda_{\min} \le \alpha) &&= \mathbb{P}(\exists \mathbf{x} : \|\mathbf{x}\|_2 = 1, \langle \mathbf{x}, W\mathbf{x}\rangle \le \alpha), \\ P_{\max}(\alpha) &= \mathbb{P}(\lambda_{\max} \ge \alpha) &&= \mathbb{P}(\exists \mathbf{x} : \|\mathbf{x}\|_2 = 1, \langle \mathbf{x}, W\mathbf{x}\rangle \ge \alpha). \end{aligned} \tag{2}$$

We use that the above is the probability of a union of events, and bound this probability from below by considering only one $\mathbf{x}$, and from above by summing over all $\mathbf{x}$. Since there are uncountably many possible $\mathbf{x}$, we will do this approximately by summing over a finite number of vectors. The lower bound for the probability yields an upper bound for the rate function, and vice versa.

We first heuristically explain the form of the rate function of $\lambda_{\max}$ and $\lambda_{\min}$, and highlight the proof. The special form of a sample covariance matrix allows us to rewrite

$$\langle \mathbf{x}, W\mathbf{x}\rangle = \frac{1}{n}\|C^T\mathbf{x}\|_2^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{m=1}^{k} x_m C_{mi}\right)^2 = \frac{1}{n}\sum_{i=1}^{n} S_{\mathbf{x},i}^2, \tag{3}$$

where

$$S_{\mathbf{x},i} = \sum_{m=1}^{k} x_m C_{mi}, \tag{4}$$

with $S_{\mathbf{x},i}$ i.i.d. for $i = 1, \dots, m$. Define

$$I_k(\alpha) = \inf_{\mathbf{x}\in\mathbb{R}^k:\|\mathbf{x}\|_2=1} \sup_t \left(t\alpha - \log \mathbb{E}[e^{tS_{\mathbf{x},1}^2}]\right). \tag{5}$$

Since $\mathbb{E}[S_{\mathbf{x},1}^2] = 1$, and $t \mapsto \log \mathbb{E}[e^{tS_{\mathbf{x},1}^2}]$ is increasing and convex, we see that, for fixed $\mathbf{x}$, the optimal t is non-negative for $\alpha \geq 1$ and non-positive for $\alpha \leq 1$. The sign of t will play an important role in the proofs in Sections 3–4.

We can now state the first result of this paper.

Theorem 2.1 *Assume that (1) holds. Then,*
(a) for all $\alpha \geq 1$ and fixed $k \geq 2$

$$\limsup_{n\to\infty} -\frac{1}{n}\log \mathbb{P}(\lambda_{\max} \geq \alpha) \leq I_k(\alpha), \tag{6}$$

and

$$\liminf_{n\to\infty} -\frac{1}{n}\log \mathbb{P}(\lambda_{\max} \geq \alpha) \geq \lim_{\varepsilon\downarrow 0} I_k(\alpha-\varepsilon), \tag{7}$$

(b) for all $0 \leq \alpha \leq 1$ and fixed $k \geq 2$

$$\limsup_{n\to\infty} -\frac{1}{n}\log \mathbb{P}(\lambda_{\min} \leq \alpha) \leq I_k(\alpha), \tag{8}$$

and

$$\liminf_{n\to\infty} -\frac{1}{n}\log \mathbb{P}(\lambda_{\min} \leq \alpha) \geq \lim_{\varepsilon\downarrow 0} I_k(\alpha+\varepsilon). \tag{9}$$

When there exists an $\epsilon > 0$ such that $\mathbb{E}[e^{\epsilon C_{11}^2}] < \infty$ and when $\mathrm{Var}(C_{11}^2) > 0$, then $I_k(\alpha) > 0$ for all $\alpha \neq 1$.

We will now discuss the main result in Theorem 2.1. Theorem 2.1 is only useful when $I_k(\alpha) > 0$, which we prove under the strong condition that there exists an $\epsilon > 0$ such that $\mathbb{E}[e^{\epsilon C_{11}^2}] < \infty$. For example, a.s. limits for the largest and smallest eigenvalues are proved under the *much* weaker condition that the fourth moment of the matrix entries C_{im} is finite. However, it is well known that the exponential bounds present in large deviations are only valid when the random variables under consideration have finite exponential moments (see e.g., Theorem 2.2 below). In this case, the rate functions can be equal to zero, and the large deviation results are rather uninformative. Since the eigenvalues are *quadratic* in the entries $\{C_{im}\}_{i,m}$, this translates into the above condition, which we therefore believe to be *necessary.*

Secondly, we note that, due to the occurrence of an infimum over $\mathbf{x}$ and a supremum over t, it is unclear whether the function $\alpha \mapsto I_k(\alpha)$ is continuous. Clearly, when $\alpha \mapsto I_k(\alpha)$ is continuous, the upper and lower bounds in (6) and (7), as well as the ones in (8) and (9), are equal. We will see that this is the case for Wishart matrices in Section 2.3. The function $\alpha \mapsto I_k(\alpha)$ can easily be seen to be increasing on $[1,\infty)$ and decreasing on $(0,1]$, since $\alpha \mapsto \sup_t \left(t\alpha - \log \mathbb{E}[e^{tS_{\mathbf{x},1}^2}]\right)$ has the same monotonicity properties for every fixed $\mathbf{x}$, so that the limits $\lim_{\varepsilon\downarrow 0} I_k(\alpha+\varepsilon)$ and $\lim_{\varepsilon\downarrow 0} I_k(\alpha-\varepsilon)$ exist as monotone limits. The continuity of $\alpha \mapsto I_k(\alpha)$ is not obvious. For example, in the simplest case where $C_{ij} = \pm 1$ with equal probability, we know that the large deviation rate function is *not* continuous, since the largest eigenvalue is at most k. Therefore, $\mathbb{P}(\lambda_{\max} \geq \alpha) = 0$ for any $\alpha > k$,

and, if $\alpha \mapsto I_k(\alpha)$ is the rate function of $\lambda_{\max}$ for $\alpha \geq 1$, then $I_k(\alpha) = \infty$ for $\alpha > k$. It remains an interesting problem to determine in what cases $\alpha \mapsto I_k(\alpha)$ *is* continuous.

Finally, we only prove that $I_k(\alpha) > 0$ for all $\alpha \neq 1$ when $\mathrm{Var}(C_{11}^2) > 0$. By the normalization that $\mathbb{E}[C_{11}] = 0, \mathbb{E}[C_{11}^2] = 1$, this only excludes the case where $C_{11} = \pm 1$ with equal probability. This case will be investigated in more detail in Theorem 4.1, where we shall also prove a lower bound implying that $I_k(\alpha) > 0$ for all $\alpha \neq 1$.

Denote

$$I_k(\alpha, \beta) = \inf_{\substack{\mathbf{x},\mathbf{y}\in\mathbb{R}^k : \|\mathbf{x}\|_2 = \|\mathbf{y}\|_2 = 1 \\ \langle \mathbf{x},\mathbf{y}\rangle = 0}} \sup_{s,t} \left(t\alpha + s\beta - \log \mathbb{E}[e^{tS_{\mathbf{x},1}^2 + sS_{\mathbf{y},1}^2}] \right). \tag{10}$$

Our proof also reveals that, and for all $0 \leq \beta \leq 1, \alpha \geq 1$,

$$\limsup_{n\to\infty} -\frac{1}{n} \log \mathbb{P}(\lambda_{\max} \geq \alpha, \lambda_{\min} \leq \beta) \geq I_k(\alpha, \beta), \tag{11}$$

and

$$\lim_{n\to\infty} -\frac{1}{n} \log \mathbb{P}(\lambda_{\max} \geq \alpha, \lambda_{\min} \leq \beta) \leq \lim_{\varepsilon \downarrow 0} I_k(\alpha + \varepsilon, \beta - \varepsilon). \tag{12}$$

For Wishart matrices, for which the entries of C are i.i.d. standard normal, the random variable $S_{\mathbf{x},i}$ has a standard normal distribution, so that we can explicitly calculate $I_k(\alpha)$. We will elaborate on this in Section 2.3 below. For the case that $C_{mi} = \pm 1$ with equal probabilities, Theorem 2.1 and its proof have also appeared in [7].

2.2 Proof of Theorem 2.1(a) and (b)

In the proof, we will repeatedly make use of the *largest-exponent-wins principle*. We first give a short explanation of this principle. This principle is about the exponential rate of the sum of two (or more) probabilities. From this point, we will abbreviate 'exponential rate of a probability' by 'rate'. Because of the minus sign, a smaller rate I means a larger exponent, and thus a larger probability. Thus, if for two events E_1 and E_2, both depending on some parameter n, we have

$$\mathbb{P}(E_1) \sim e^{-nI_1} \quad \text{and} \quad \mathbb{P}(E_2) \sim e^{-nI_2}$$

then

$$-\lim_{n\to\infty} \frac{1}{n} \log(\mathbb{P}(E_1) + \mathbb{P}(E_2)) \sim \min\{I_1, I_2\}. \tag{13}$$

In words, the principle states that as $n \to \infty$, the smallest exponent (i.e., the largest rate) will become negligible. It also implies that

$$-\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(E_1 \cup E_2) \sim \min\{I_1, I_2\}. \tag{14}$$

In the proof, we will make essential use of Cramér's Theorem, which we state here for the sake of completeness:

Theorem 2.2 (Cramér's theorem and Chernoff bound) *Let $(X_i)_{i=1}^\infty$ be a sequence of i.i.d. random variables. Then, for all $a \geq \mathbb{E}[X_1]$,*

$$-\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\Big(\frac{1}{n}\sum_{i=1}^n X_i \geq a\Big) = \sup_{t\geq 0} \left(ta - \log \mathbb{E}[e^{tX_1}] \right), \tag{15}$$

while, for all $a \leq \mathbb{E}[X_1]$,

$$-\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\Big(\frac{1}{n}\sum_{i=1}^n X_i \leq a\Big) = \sup_{t\leq 0} \left(ta - \log \mathbb{E}[e^{tX_1}] \right). \tag{16}$$

The upper bounds in (15)-(16) hold for every n.
Furthermore, when $\mathbb{E}[e^{tX_1}] < \infty$ for all t with $|t| \leq \epsilon$ and some $\epsilon > 0$, then the right-hand sides of (15) and (16) are strictly positive for all $a \neq \mathbb{E}[X_1]$.

See e.g., [20, Theorem 1.1, pages 5-6 and Proposition 1.9, page 13] for this result, and see [6] and [12] for general introductions to large deviation theory.

For the proof, we start by showing that $I_k(\alpha) > 0$ for all $\alpha \neq 1$ when there exists an $\epsilon > 0$ such that $\mathbb{E}[e^{\epsilon C_{11}^2}] < \infty$ and when $\mathrm{Var}(C_{11}^2) > 0$. For this, we note that, by the Cauchy-Schwarz inequality and (4), for every $\mathbf{x}$ with $\|\mathbf{x}\|_2 = 1$,

$$S_{\mathbf{x},i}^2 \leq \sum_{m=1}^k x_m^2 \sum_{m=1}^k C_{mi}^2 = \sum_{m=1}^k C_{mi}^2,$$

so that $\mathbb{E}[e^{tS_{\mathbf{x},i}^2}] \leq \mathbb{E}[e^{tC_{11}^2}]^k < \infty$ whenever there exists an $\epsilon > 0$ such that $\mathbb{E}[e^{\epsilon C_{11}^2}] < \infty$. Thus, uniformly in $\mathbf{x}$ with $\|\mathbf{x}\|_2 = 1$, the random variables $S_{\mathbf{x},i}^2$ have bounded exponential moments for $t \leq \epsilon$. As a result, the Taylor expansion

$$\log \mathbb{E}[e^{tS_{\mathbf{x},i}^2}] = t + \frac{t^2}{2}\mathrm{Var}(S_{\mathbf{x},i}^2) + \mathcal{O}(|t|^3) \tag{17}$$

holds uniformly in $\mathbf{x}$ with $\|\mathbf{x}\|_2 = 1$. We compute, since $\mathbb{E}[S_{\mathbf{x},i}^2] = \mathbb{E}[C_{11}^2] = 1$, and for $\mathbf{x}$ with $\|\mathbf{x}\|_2 = 1$,

$$\mathbb{E}[S_{\mathbf{x},i}^4] = 3\Big(\sum_m x_m^2\Big)^2 - 3\sum_m x_m^4 + \mathbb{E}[C_{11}^4]\sum_m x_m^4 = 3 - 3\sum_m x_m^4 + \mathbb{E}[C_{11}^4]\sum_m x_m^4,$$

that

$$\mathrm{Var}(S_{\mathbf{x},i}^2) = 3 - 3\sum_m x_m^4 + \mathbb{E}[C_{11}^4]\sum_m x_m^4 - 1 = 2 - 2\sum_m x_m^4 + \mathrm{Var}(C_{11}^2)\sum_m x_m^4,$$

which is bounded, since by assumption $\mathbb{E}[e^{tC_{11}^2}] < \infty$. Furthermore, $\sum_m x_m^4 \in [0,1]$ uniformly in $\mathbf{x}$ with $\|\mathbf{x}\|_2 = 1$, so that, again uniformly in $\mathbf{x}$ with $\|\mathbf{x}\|_2 = 1$, $\mathrm{Var}(S_{\mathbf{x},i}^2) \geq \min\{2, \mathrm{Var}(C_{11}^2)\} > 0$. We conclude that, for t sufficiently small, uniformly in $\mathbf{x}$ with $\|\mathbf{x}\|_2 = 1$, and by ignoring higher-order Taylor expansion terms of $t \mapsto \log \mathbb{E}[e^{tS_{\mathbf{x},i}^2}]$ in (17), which is allowed when $|t|$ is sufficiently small,

$$\log \mathbb{E}[e^{tS_{\mathbf{x},i}^2}] \leq t + t^2 \min\{2, \mathrm{Var}(C_{11}^2)\}.$$

In turn, this implies that for $|t| \leq \epsilon$ small, and uniformly in $\mathbf{x}$ with $\|\mathbf{x}\|_2 = 1$,

$$I_k(\alpha) \geq \inf_{\mathbf{x}\in\mathbb{R}^k:\|\mathbf{x}\|_2=1} \sup_{|t|\leq\epsilon} \Big(t\alpha - \log \mathbb{E}[e^{tS_{\mathbf{x},1}^2}]\Big) \geq \inf_{\mathbf{x}\in\mathbb{R}^k:\|\mathbf{x}\|_2=1} \sup_{|t|\leq\epsilon} \left(t(\alpha-1) - \frac{t^2}{2}\min\{2, \mathrm{Var}(C_{11}^2)\}\right) > 0,$$

the latter bound holding for every $\alpha \neq 1$ when $\mathrm{Var}(C_{11}^2) > 0$. This completes the proof that $I_k(\alpha) > 0$ for all $\alpha \neq 1$ when there exists an $\epsilon > 0$ such that $\mathbb{E}[e^{\epsilon C_{11}^2}] < \infty$ and $\mathrm{Var}(C_{11}^2) > 0$.

We continue by proving (6)–(9). The proof for $\lambda_{\max}$ is similar to the one for $\lambda_{\min}$, so we will focus on the latter. To obtain the upper bound of the rate of (3), we use that for any $\mathbf{x}'$ with $\|\mathbf{x}'\|_2 = 1$,

$$\mathbb{P}(\lambda_{\min} \leq \alpha) = \mathbb{P}(\exists \mathbf{x} : \langle \mathbf{x}, W\mathbf{x}\rangle \leq \alpha) \geq \mathbb{P}(\langle \mathbf{x}', W\mathbf{x}'\rangle \leq \alpha). \tag{18}$$

Now insert (3). Since $\mathbf{x}'$ is fixed, the $S_{\mathbf{x}',i}^2$ are i.i.d. variables, and we can apply Cramér's Theorem to obtain the upper bound for the rate function for fixed $\mathbf{x}'$. This yields that, for every $\mathbf{x}'$, we have

$$-\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\lambda_{\min} \leq \alpha) \leq \sup_t \Big(t\alpha - \log \mathbb{E}[e^{tS_{\mathbf{x}',1}^2}]\Big). \tag{19}$$

If we maximize the right hand side over $\mathbf{x}'$, then we arrive at $I_k(\alpha)$ as the upper bound, and we have proved (8). The proof for (6) is identical.

We are left to prove the lower bounds in (7) and (9). For this, we wish to sum over all possible $\mathbf{x}$. We approximate the sphere $\|\mathbf{x}\|_2 = 1$ by a finite set of vectors $\mathbf{x}^{(j)}$ with $\|\mathbf{x}^{(j)}\|_2 = 1$, such that the distance between two of these vectors is at most d, and observe that

$$\begin{aligned}
|\langle \mathbf{x}, W\mathbf{x}\rangle - \langle \mathbf{x}^{(j)}, W\mathbf{x}^{(j)}\rangle| &= |\langle (\mathbf{x}-\mathbf{x}^{(j)}), W\mathbf{x}\rangle + \langle \mathbf{x}^{(j)}, W(\mathbf{x}-\mathbf{x}^{(j)})\rangle| \\
&= |\langle \mathbf{x}, W(\mathbf{x}-\mathbf{x}^{(j)})\rangle + \langle \mathbf{x}^{(j)}, W(\mathbf{x}-\mathbf{x}^{(j)})\rangle| \\
&\leq (\|\mathbf{x}\| + \|\mathbf{x}^{(j)}\|) \cdot \|W\| \cdot \|\mathbf{x}-\mathbf{x}^{(j)}\| \leq 2\lambda_{\max} d.
\end{aligned}$$

We need that $\lambda_{\max} \leq \kappa k$, with κ some large enough constant, with sufficiently high probability, which we will prove first. We have that $\lambda_{\max} \leq T_W$, where T_W is the trace of W, since W is non-negative. Note that

$$T_W = \frac{1}{n} \sum_{i=1}^{n} \sum_{m=1}^{k} C_{mi}^2. \tag{20}$$

Thus, T_W is a sum of nk i.i.d. variables.

Since $\mathbb{E}[e^{tC_{11}^2}] < \infty$ for all $t \leq \epsilon$, we can use Cramér's Theorem for T_W. Therefore, for any κ, by the Chernoff bound,

$$\mathbb{P}(T_W > \kappa k) \leq e^{-nkI_{C^2}(\kappa)}, \tag{21}$$

where

$$I_{C^2}(a) = \sup_t \left(ta - \log \mathbb{E}[e^{tC_{11}^2}] \right). \tag{22}$$

Since $\mathbb{E}[C_{11}^2] = \mathrm{Var}(C_{11}) = 1$, we have that $I_{C^2}(\kappa) > 0$ for any $\kappa > 1$. Therefore, by picking $\kappa > 1$ large enough, we can make $kI_{C^2}(\kappa)$ arbitrarily large. If we take $kI_{C^2}(\kappa)$ larger than $I_k(\alpha - \varepsilon)$, according to (13), this will not influence the result. (Note that when $I_k(\alpha - \varepsilon) = \infty$ for all $\varepsilon > 0$, then we can also let $kI_{C^2}(\kappa)$ tend to infinity by taking $\kappa \to \infty$.)

It follows that

$$\begin{aligned}
\mathbb{P}(\lambda_{\min} \leq \alpha) &\leq \mathbb{P}(\exists \mathbf{x}^{(j)} : \langle \mathbf{x}^{(j)}, W\mathbf{x}^{(j)} \rangle \leq \alpha + 2d\kappa k) + \mathbb{P}(T_W > \kappa k) \\
&\leq \sum_j \mathbb{P}(\langle \mathbf{x}^{(j)}, W\mathbf{x}^{(j)} \rangle \leq \alpha + 2d\kappa k) + \mathbb{P}(T_W > \kappa k) \\
&\leq N_d \sup_{\mathbf{x}^{(j)}} \mathbb{P}(\langle \mathbf{x}^{(j)}, W\mathbf{x}^{(j)} \rangle \leq \alpha + 2d\kappa k) + \mathbb{P}(T_W > \kappa k),
\end{aligned} \tag{23}$$

with N_d the number of vectors in the finite approximation of the sphere. The above bound is valid for every choice of κ, k, α and d.

We write $\varepsilon = 2d\kappa k$ and will later let $\varepsilon \downarrow 0$. Then, applying the largest-exponent-wins principle for $\kappa > 0$ large enough, as well as Cramér's Theorem together with (3), we arrive at

$$\begin{aligned}
-\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\lambda_{\min} \leq \alpha) &\geq \inf_{\mathbf{x}^{(j)}} \sup_t \left(t(\alpha + \varepsilon) - \log \mathbb{E}[e^{tS^2_{\mathbf{x}^{(\mathrm{j})},1}}] \right) + \liminf_{n\to\infty} \frac{1}{n} \log N_d \\
&\geq I_k(\alpha + \varepsilon) + \liminf_{n\to\infty} \frac{1}{n} \log N_d.
\end{aligned} \tag{24}$$

In a similar way, we obtain that

$$\begin{aligned}
-\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\lambda_{\max} \geq \alpha) &\geq \inf_{\mathbf{x}^{(j)}} \sup_t \left(t(\alpha - \varepsilon) - \log \mathbb{E}[e^{tS^2_{\mathbf{x}^{(\mathrm{j})},1}}] \right) + \liminf_{n\to\infty} \frac{1}{n} \log N_d \\
&\geq I_k(\alpha - \varepsilon) + \liminf_{n\to\infty} \frac{1}{n} \log N_d,
\end{aligned} \tag{25}$$

where we take d so small that $\alpha - \varepsilon > 0$.

A simple overestimation of N_d is obtained by first taking $[-1, 1]^k \subset \mathbb{R}^k$ around the origin, and laying a grid on this cube with grid length $\frac{1}{L}$. We then normalize the centers of these cubes to have norm 1. The finite set of vectors consists of the centers of the small cubes of width $2/L$. In this case,

$$d \leq \frac{3\sqrt{k}}{L}, \qquad \text{and} \qquad N_d \leq L^k. \tag{26}$$

Indeed, the first bound follows since, for any vector $\mathbf{x}$, there exists a center of a small cube for which all coordinates are at most $1/L$ away. Therefore, the distance of to this center is at most $\frac{\sqrt{k}}{L}$. Since $\mathbf{x}$ has norm 1, the norm of the center of the cube is in between $1 - \frac{\sqrt{k}}{L}$ and $1 + \frac{\sqrt{k}}{L}$, and we obtain that the distance of $\mathbf{x}$ to the normalized center of the small cube is at most

$$d \leq \frac{\sqrt{k}}{L} + \frac{\frac{\sqrt{k}}{L}}{1 - \frac{\sqrt{k}}{L}} \leq 3\frac{\sqrt{k}}{L}, \tag{27}$$

when $\frac{\sqrt{k}}{L} \leq 1/2$. For this choice, we have $\varepsilon = 6\kappa k^{3/2}/L$, which we can make small by taking L large.

We conclude that, for any $L < \infty$, $\lim_{n\to\infty} \frac{1}{n} \log N_d = 0$, so that, for any $\kappa > 1$ sufficiently large,

$$-\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\lambda_{\min} \leq \alpha) \geq I_k(\alpha + \varepsilon), \tag{28}$$

and

$$-\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\lambda_{\max} \geq \alpha) \geq I_k(\alpha - \varepsilon), \tag{29}$$

when the respective right-hand sides are finite. Since the above statement is true for any ε, we can take $\varepsilon \downarrow 0$ by letting $L \uparrow \infty$. When the right-hand side are infinite, then we conclude that also the left-hand sides can be made arbitrarily large by letting $L \uparrow \infty$. This completes the proof of (7) and (9). □

To see (11)–(12), we follow the above proof. We first note that the eigenvectors corresponding to $\lambda_{\max}$ and $\lambda_{\min}$ are orthogonal. Therefore, we obtain that

$$\mathbb{P}(\lambda_{\max} \geq \alpha, \lambda_{\min} \leq \beta) = \mathbb{P}\big(\exists \mathbf{x}, \mathbf{y} : \|\mathbf{x}\|_2 = \|\mathbf{y}\|_2 = 1, \langle \mathbf{x}, \mathbf{y}\rangle = 0, \langle \mathbf{x}, W\mathbf{x}\rangle \geq \alpha, \langle \mathbf{y}, W\mathbf{y}\rangle \leq \beta\big). \tag{30}$$

We now proceed as above, and for the lower bound pick any $\mathbf{x}, \mathbf{y}$ satisfying the requirements in the probability on the right hand side. The upper bound is slightly harder. For this, we need to pick a finite approximation for the choices of $\mathbf{x}$ and $\mathbf{y}$ such that $\|\mathbf{x}\|_2 = \|\mathbf{y}\|_2 = 1$ and $\langle \mathbf{x}, \mathbf{y}\rangle = 0$. We will now show that we can do this in such a way that the total number of pairs $\{\mathbf{x}^{(i)}, \mathbf{y}^{(i,j)}\}_{i,j\geq 1}$ is bounded by N_d^2, where N_d is as in (26).

We pick $\{\mathbf{x}^{(i)}\}_{i\geq 1}$ as in the above proof. Then, for fixed $\mathbf{x}^{(i)}$, we define a finite number of $\mathbf{y}$ such that $\langle \mathbf{x}^{(i)}, \mathbf{y}\rangle = 0$. For this, we consider, for fixed $\mathbf{x}^{(i)}$, only those cubes of width $\frac{1}{L}$ around an $\mathbf{x}^{(j)}$, for some j, that contain at least one element $\mathbf{z}$ having norm 1 and such that $\langle \mathbf{z}, \mathbf{x}^{(i)}\rangle = 0$. Fix one of such cubes. If there are more such $\mathbf{z}$ in this cube around $\mathbf{x}^{(j)}$, then we pick the unique element that is closest to $\mathbf{x}^{(j)}$. We denote this element by $\mathbf{y}^{(j,i)}$. The set of these elements $\mathbf{y}^{(j,i)}$ will be denoted by $\{\mathbf{y}^{(j,i)}\}_{i\geq 1}$. The finite subset of the set $\|\mathbf{x}\|_2 = \|\mathbf{y}\|_2 = 1$ and $\langle \mathbf{x}, \mathbf{y}\rangle = 0$ then consists of $\{\mathbf{x}^{(i)}, \mathbf{y}^{(i,j)}\}_{i,j\geq 1}$.

We clearly have that every $\mathbf{x}$ and $\mathbf{y}$ with $\|\mathbf{x}\|_2 = \|\mathbf{y}\|_2 = 1$ and $\langle \mathbf{x}, \mathbf{y}\rangle = 0$ can be approximated by a pair $\mathbf{x}^{(j)}$ and $\mathbf{y}^{(j,i)}$ such that $\|\mathbf{x} - \mathbf{x}^{(j)}\|_2 \leq d$ and $\|\mathbf{y} - \mathbf{y}^{(i,j)}\|_2 \leq 2d$. Then we can complete the proof as above.

2.3 Special case: Wishart matrices

To give an example, we go to Wishart matrices, for which C_{ij} are i.i.d. standard normal. In this case, we can compute $I_k(\alpha)$ and $I_k(\alpha, \beta)$ explicitly. To compute $I_k(\alpha)$, we note that, for any $\mathbf{x}$ such that $\|\mathbf{x}\|_2 = 1$, we have that $S_{\mathbf{x},1}$ is standard normal. Therefore,

$$\mathbb{E}[e^{tS_{\mathbf{x},1}^2}] = \frac{1}{\sqrt{1-2t}}, \tag{31}$$

so that

$$I_k(\alpha) = \sup_t \Big(t\alpha - \log\Big(\frac{1}{\sqrt{1-2t}}\Big)\Big). \tag{32}$$

In order to compute $I_k(\alpha)$, we note that the maximization problem over t in $\sup_t t\alpha - \log\big(\frac{1}{\sqrt{1-2t}}\big)$ is straightforward, and yields $t^* = \frac{1}{2} - \frac{1}{2\alpha}$ and $I_k(\alpha) = \frac{1}{2}(\alpha - 1 - \log\alpha)$. Note that $I_k(\alpha)$ is independent of k. In particular, we see that $\alpha \mapsto I_k(\alpha)$ is continuous, which leads us to the following corollary:

Corollary 2.3 *Let C_{ij} be independent standard normals. Then,*
(a) for all $\alpha \geq 1$ and fixed $k \geq 2$

$$\lim_{n\to\infty} -\frac{1}{n} \log \mathbb{P}(\lambda_{\max} \geq \alpha) = \frac{1}{2}(\alpha - 1 - \log\alpha), \tag{33}$$

(b) for all $0 \leq \alpha \leq 1$ and fixed $k \geq 2$

$$\limsup_{n\to\infty} -\frac{1}{n} \log \mathbb{P}(\lambda_{\min} \leq \alpha) = \frac{1}{2}(\alpha - 1 - \log\alpha). \tag{34}$$

We next turn to the computation of $I_k(\alpha,\beta)$. When $\mathbf{x}$ and $\mathbf{y}$ are such that $\|\mathbf{x}\|_2 = \|\mathbf{y}\|_2 = 1$ and $\langle \mathbf{x}, \mathbf{y}\rangle = 0$, then we have that $(S_{\mathbf{x},1}, S_{\mathbf{y},1})$ are normally distributed. It can easily be seen that $\mathbb{E}[S_{\mathbf{x},1}] = 0, \mathbb{E}[S_{\mathbf{x},1}^2] = \|\mathbf{x}\|_2^2 = 1$, so that $S_{\mathbf{x},1}$ and $S_{\mathbf{y},1}$ are standard normal. Moreover, $\mathbb{E}[S_{\mathbf{x},1}S_{\mathbf{y},1}] = \langle \mathbf{x}, \mathbf{y}\rangle = 0$, so that $(S_{\mathbf{x},1}, S_{\mathbf{y},1})$ are in fact independent standard normal random variables. Therefore,

$$\mathbb{E}[e^{tS_{\mathbf{x},1}^2 + sS_{\mathbf{y},1}^2}] = \mathbb{E}[e^{tS_{\mathbf{x},1}^2}]\mathbb{E}[e^{sS_{\mathbf{y},1}^2}] = \frac{1}{\sqrt{1-2t}}\frac{1}{\sqrt{1-2s}}, \tag{35}$$

and, for $\alpha \in [0,1]$ and $\beta \geq 1$,

$$I_k(\alpha,\beta) = \sup_{s,t}\left(t\alpha + s\beta - \log\left(\frac{1}{\sqrt{1-2t}}\right) - \log\left(\frac{1}{\sqrt{1-2s}}\right)\right) = I_k(\alpha) + I_k(\beta), \tag{36}$$

so that the exponential rate of the probability that $\lambda_{\max} \geq \alpha$ and $\lambda_{\min} \leq \beta$ is the exponential rate of the product of the probabilities that $\lambda_{\max} \geq \alpha$ and $\lambda_{\min} \leq \beta$. This remarkable form of independence seems to be true only for Wishart matrices.

The above considerations lead to the following corollary:

Corollary 2.4 *Let C_{ij} be independent standard normals. Then,*

$$\lim_{n\to\infty} -\frac{1}{n}\log \mathbb{P}(\lambda_{\max} \geq \alpha, \lambda_{\min} \leq \beta) = \frac{1}{2}(\alpha - 1 - \log\alpha) + \frac{1}{2}(\beta - 1 - \log\beta). \tag{37}$$

In the sequel, we will, among other things, investigate cases where, for $k \to \infty$, the rate function $I_k(\alpha)$ for general C_{ij} converges to the Gaussian limit $I_\infty(\alpha) = \frac{1}{2}(\alpha - 1 - \log\alpha)$.

3 Asymptotics for the eigenvalues for symmetric and bounded entries of C

In this section, we investigate the case where C_{mi} is symmetric around 0 and $|C_{mi}| < M < \infty$ almost surely, or C_{mi} is standard normal. To emphasize the role of k, we will denote the law of W for a given k by $\mathbb{P}_k$. We define the extension to $k = \infty$ of $I_k(\alpha)$ to be

$$I_\infty(\alpha) = \inf_{\mathbf{x}\in\ell^2(\mathbb{N}):\|\mathbf{x}\|_2=1} \sup_t \left(t\alpha - \log\mathbb{E}[e^{tS_{\mathbf{x},1}^2}]\right), \tag{38}$$

where $\ell^2(\mathbb{N})$ is the space of all infinite square-summable sequences, with norm $\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^\infty \mathbf{x}_i^2}$. The main result in this section is the following theorem:

Theorem 3.1 *Suppose that C_{mi} is symmetric around zero and that $|C_{mi}| < M < \infty$ almost surely, or C_{mi} is standard normal. Then, for all $k_n \to \infty$ such that $k_n = o(\frac{n}{\log\log n})$,*
(a) for all $\alpha \geq 1$,

$$\liminf_{n\to\infty} -\frac{1}{n}\log\mathbb{P}_{k_n}(\lambda_{\max} \geq \alpha) \leq I_\infty(\alpha), \tag{39}$$

and

$$\limsup_{n\to\infty} -\frac{1}{n}\log\mathbb{P}_{k_n}(\lambda_{\max} \geq \alpha) \geq \lim_{\varepsilon\downarrow 0} I_\infty(\alpha - \varepsilon), \tag{40}$$

(b) for all $0 < \alpha \leq 1$,

$$\liminf_{n\to\infty} -\frac{1}{n}\log\mathbb{P}_{k_n}(\lambda_{\min} \leq \alpha) \leq I_\infty(\alpha), \tag{41}$$

and

$$\liminf_{n\to\infty} -\frac{1}{n}\log\mathbb{P}_{k_n}(\lambda_{\min} \leq \alpha) \geq \lim_{\varepsilon\downarrow 0} I_\infty(\alpha + \varepsilon). \tag{42}$$

A version of this result has also been published in a conference proceeding [7], for the special case $C_{mi} = \pm 1$, each with probability 1/2, and where the restriction on k_n was $k_n = \mathcal{O}(\frac{n}{\log n})$. Unfortunately, there is a technical error in the proof, and below we present the corrected proof. In order to do so, we will rely on explicit lower bounds for $I_k(\alpha)$ for $\alpha \geq 1$.

A priory, it is not obvious that the limit $I_\infty(\alpha)$ is strictly positive for $\alpha \neq 1$. However, in the examples we will investigate later on, such as $C_{mi} = \pm 1$ with equal probability, we will see that indeed $I_\infty(\alpha) > 0$ for $\alpha \neq 1$. Possibly, such a result can be shown more generally.

The following proposition is instrumental in the proof of Theorem 3.1:

Proposition 3.2 *Assume that C_{mi} is symmetric around zero and that $|C_{mi}| < M < \infty$ almost surely, or C_{mi} is standard normal. Then, for all k, $\alpha \geq M^2$ and $\mathbf{x}$ with $\|\mathbf{x}\|_2 = 1$,*

$$\mathbb{P}_k(\langle \mathbf{x}, W\mathbf{x}\rangle \geq \alpha) \leq e^{-nJ_k(\alpha)}, \tag{43}$$

where

$$J_k(\alpha) = \frac{1}{2}\Big(\frac{\alpha}{M^2} - 1 - \log\frac{\alpha}{M^2}\Big). \tag{44}$$

In the case where $C_{mi} = \pm 1$, for which $M > 1$, we will present an improved version of this bound, valid when $\alpha \geq 1/2$, in Theorem 4.1 below.

3.1 Proof of Proposition 3.2

Throughout this proof, we fix $\mathbf{x}$ with $\|x\|_2 = 1$. We use (3) to bound, for every $t \geq 0$ and $k \in \mathbb{N}$, by the Markov inequality,

$$\mathbb{P}_k(\langle \mathbf{x}, W\mathbf{x}\rangle \geq \alpha) = \mathbb{P}_k(e^{t\sum_{i=1}^n S_{\mathbf{x},i}^2} \geq e^{nt\alpha}) \leq e^{-n\left(\alpha t - \log \mathbb{E}_{k_n}[e^{tS_{\mathbf{x},1}^2}]\right)}. \tag{45}$$

We claim that for all $0 \leq t \leq \frac{1}{M^2}$,

$$\mathbb{E}_{k_n}[e^{tS_{\mathbf{x},i}^2}] \leq \frac{1}{\sqrt{1 - 2M^2 t}}. \tag{46}$$

In the case of Wishart matrices, for which $S_{\mathbf{x},i}$ has a standard normal distribution, (46) holds with equality for $M = 1$.

We first note that (46) is proven in [15, Section IV], for the case that $C_{ij} = \pm 1$ with equal probability. For any k and $\mathbf{x}$, the bound is even valid for all $-1/2 \leq t \leq 1/2$. We now extend the case where $C_{ij} = \pm 1$ to the case where C_{ij} is symmetric around zero and satisfies $|C_{ij}| < M$ almost surely.

We write $C_{ij} = A_{ij}C_{ij}^*$, where $A_{ij} = |C_{ij}| < M$ a.s. and $C_{ij}^* = \mathrm{sign}(C_{ij})$. Moreover, A_{ij} and C_{ij}^* are independent, since C_{ij} has a symmetric distribution around zero. Thus, we obtain that $S_{\mathbf{x},i} = S^*_{A_i\mathbf{x},i}$, where $(A_i\mathbf{x})_j = A_{ij}\mathbf{x}_j$, and

$$S^*_{\mathbf{y},i} = \sum_{j=1}^k C_{ij}^* \mathbf{y}_j. \tag{47}$$

For $S^*_{\mathbf{y},i}$ we know that (46) is proven. Therefore,

$$\mathbb{E}_k[e^{tS^2_{A_i\mathbf{x},i}}] \leq \mathbb{E}_k\Big[\frac{1}{\sqrt{1 - 2t\|A_i\mathbf{x}\|_2^2}}\Big] \tag{48}$$

for all t such that $-1/2 \leq t\|A_i\mathbf{x}\|_2 \leq 1/2$ almost surely. When $\|\mathbf{x}\|_2^2 = 1$, we have that

$$0 \leq \|A_i\mathbf{x}\|_2 < M \qquad \text{a.s.} \tag{49}$$

Therefore, $\mathbb{E}_k[e^{tS^2_{A_i\mathbf{x},i}}] \leq \frac{1}{\sqrt{1 - 2M^2 t\|\mathbf{x}\|_2^2}}$ for all $0 \leq tM^2\|\mathbf{x}\|_2^2 \leq 1/2$. Thus, we arrive at

$$\mathbb{P}_k(\langle \mathbf{x}, W\mathbf{x}\rangle \geq \alpha) \leq e^{-n\left(\sup_{0\leq t\leq 1/M^2}\left(t\alpha - \log\frac{1}{\sqrt{1-2M^2t}}\right)\right)}. \tag{50}$$

Note that since $\|\mathbf{x}\|_2 = 1$, the bound is independent of $\mathbf{x}$. performing the maximum over t on the right-hand side of (50) over t yields $t^* = \frac{1}{2M^2} - \frac{1}{2\alpha}$, and inserting this value t^* in the right-hand side gives the result. □

3.2 Proof of Theorem 3.1

The proof is similar to that of Theorem 2.1. For the proofs of (39) and (41), we again use (18), but now choose an $\mathbf{x}'$ of which only the first k components are non-zero. This leads to, using that $k_n \to \infty$, so that $k_n \geq k$ for n sufficiently large,

$$\liminf_{n\to\infty} -\frac{1}{n}\log \mathbb{P}_{k_n}(\lambda_{\max} \geq \alpha) \leq \sup_t \left(t\alpha - \log \mathbb{E}[e^{tS^2_{\mathbf{x}',1}}]\right). \tag{51}$$

Maximizing over all $\mathbf{x}'$ of which only the first k components are non-zero leads to

$$\liminf_{n\to\infty} -\frac{1}{n}\log \mathbb{P}_{k_n}(\lambda_{\max} \geq \alpha) \leq I_k(\alpha), \tag{52}$$

where this bound is valid for all $k \in \mathbb{N}$. We next claim that

$$\lim_{k\to\infty} I_k(\alpha) = I_\infty(\alpha). \tag{53}$$

For this, we first note that the sequence $k \mapsto I_k(\alpha)$ is non-increasing and non-negative, so that it has a pointwise limit. Secondly, $I_k(\alpha) \geq I_\infty(\alpha)$ for all k, since the possible choices of $\mathbf{x}$ in (38) is larger than the one in (5). Now it is not hard to see that $\lim_{k\to\infty} I_k(\alpha) = I_\infty(\alpha)$, by splitting into the two cases depending on whether the infimum over $\mathbf{x}$ in (38) is attained or not. This completes the proof of (39) and (41).

For the proof of (40) and (42), we adapt the proof (7) and (9). As in the proof of Theorem 2.1(a-b), we wish to show that the terms $\frac{1}{n}\log N_d$ and $2d\lambda_{\max}$ vanish when we take the logarithm of (23), divide by n and let $n \to \infty$. However, this time we wish to let $k_n \to \infty$ as well, for k_n as large as possible. We will have $k_n = o(n)$ in mind.

The overestimation (26) can be improved using an upper bound for the number $M_R = N_{1/R}$ of spheres of radius $1/R$ needed to cover the surface of a k-dimensional sphere of radius 1 (Rogers (1963)), when $k \to \infty$,

$$M_R = 4k\sqrt{k}R^k\left(\log k + \log\log k + \log(R)\right)\left(1 + \mathcal{O}(1/\log k)\right) \equiv f(k,R)R^k. \tag{54}$$

This bound is valid for $R > \sqrt{\frac{k}{k-1}}$. Since we use small spheres this time, $d \leq 1/R$.

We can also improve the upper bound for $\lambda_{\max}$. For any $\Omega_n > 1$, which we will choose appropriately later on, we split

$$P_{\min}(\alpha) \leq \mathbb{P}(\lambda_{\min} \leq \alpha, \lambda_{\max} \leq \Omega_n) + P_{\max}(\Omega_n), \tag{55}$$

$$P_{\max}(\alpha) = \mathbb{P}(\alpha \leq \lambda_{\max} \leq \Omega_n) + P_{\max}(\Omega_n). \tag{56}$$

We first give a sketch of the proof, omitting the details. The idea is that the first term of these expressions will yield the rate function $I_\infty(\alpha)$. The term $P_{\max}(\Omega_n)$ has an exponential rate which is $\mathcal{O}(\Omega_n) - \mathcal{O}(\frac{k_n \log k_n}{n})$, and, since $\frac{k_n \log k_n}{n} = o(\log n)$, can thus be made arbitrarily large by taking $\Omega_n = K\log n$ with $K > 1$ large enough. This means that we can choose Ω_n large enough to make this rate disappear according to the largest-exponent-wins principle (13). We will need different choices of R for the two terms. We will now give the details of the proof.

We first bound $P_{\max}(\Omega_n)$ of (56), using (23). In (23), we choose $\kappa = M^2$. This leads to

$$P_{\max}(\Omega_n) \leq M_R \sup_{\mathbf{x}} \mathbb{P}(\langle \mathbf{x}, W\mathbf{x}\rangle \geq \Omega_n - 2dM^2k_n) + \mathbb{P}(T_W \geq M^2 k_n),$$

where the supremum over $\mathbf{x}$ runs over the centers of the small balls. Inserting (54), choosing $R = k_n$ and using $d \leq 1/R$, this becomes

$$P_{\max}(\Omega_n) \leq f(k_n,k_n)k_n^{k_n} \sup_{\mathbf{x}} \mathbb{P}(\langle \mathbf{x}, W\mathbf{x}\rangle \geq \Omega_n - 2M^2) + \mathbb{P}(T_W \geq M^2 k_n).$$

Using Proposition 3.2, we find

$$\begin{aligned} P_{\max}(\Omega_n) \quad &\leq f(k_n,k_n)k_n^{k_n} e^{-\frac{1}{2}n(\frac{\Omega_n}{M^2} - 3 - \log(\frac{\Omega_n}{M^2} - 2))} + e^{-nk_n I_{C^2_{11}}(M^2)} \\ &= f(k_n,k_n)e^{k_n\log k_n - \frac{1}{2}n\left(\frac{\Omega_n}{M^2} - 3 - \log(\frac{\Omega_n}{M^2} - 2)\right)} + e^{-nk_n I_{C^2_{11}}(M^2)}. \end{aligned} \tag{57}$$

We choose $\Omega_n = K\log n$ with K so large that

$$\frac{k_n \log k_n}{n} < \frac{1}{4}\Big(\frac{\Omega_n}{M^2} - 3 - \log(\frac{\Omega_n}{M^2} - 2)\Big).$$

Therefore, also using that

$$f(k_n, k_n) = e^{o(n\log n)},$$

we obtain

$$P_{\max}(\Omega_n) \leq e^{-\frac{K}{4M^2} n\log n(1+o(1))} + e^{-nk_n I_{C_{11}^2}(M^2)}. \tag{58}$$

Next, we investigate the first term of (56). In this term, we can use Ω_n as the upper bound for $\lambda_{\max}$. Therefore, again starting with (23), we obtain that, for any R,

$$-\frac{1}{n}\log \mathbb{P}(\lambda_{\min} \leq \alpha, \lambda_{\max} \leq \Omega_n) \geq -\frac{1}{n}\big[\log M_R + \sup_{\mathbf{x}} \log \mathbb{P}\big(\langle \mathbf{x}, W\mathbf{x}\rangle \leq \alpha + 2\Omega_n/R\big)\big]. \tag{59}$$

For $\lambda_{\max}$, we get a similar expression. Inserting (54), we need to choose R again. This time we wish to choose $R = R_n$ to increase in such a way that $k_n \log R_n = o(n)$ and $\Omega_n = K\log n = o(R_n)$. For the latter, we need that $R_n \gg \log n$, so that we can only satisfy the first restriction when $k_n = o(\frac{n}{\log\log n})$. Then this is sufficient to make the term $2\Omega_n/R_n$ disappear as $k_n \to \infty$, and to make the term $\frac{1}{n}(\log M_{R_n}) = \mathcal{O}(k_n \log R_n)$ to be $o(n)$, so that it also disappears. Therefore, for any $R = R_n$ satisfying the above two restrictions,

$$-\frac{1}{n}\log \mathbb{P}(\lambda_{\min} \leq \alpha, \lambda_{\max} \leq \Omega_n) \geq I_{k_n}(\alpha + 2\Omega_n/R_n) + o(1). \tag{60}$$

Similarly,

$$-\frac{1}{n}\log \mathbb{P}(\alpha \leq \lambda_{\max} \leq \Omega_n) \geq I_{k_n}(\alpha - 2\Omega_n/R_n) + o(1). \tag{61}$$

Moreover, by the fact that $\Omega_n = K\log n = o(R_n)$, we have that $2\Omega_n/R_n \leq \varepsilon$ for all n large enough. By the monotonicity of $\alpha \mapsto I_{k_n}(\alpha)$, we then have that

$$I_{k_n}(\alpha + 2\Omega_n/R_n) \geq I_{k_n}(\alpha + \varepsilon), \qquad I_{k_n}(\alpha - 2\Omega_n/R_n) \geq I_{k_n}(\alpha - \varepsilon). \tag{62}$$

Since $\lim_{k\to\infty} I_k(\alpha) = I_\infty(\alpha)$ (see (53)), putting (60), (62) and (58) together and applying the largest-exponent-wins principle (13), we see that the proof follows when

$$I_\infty(\alpha \pm \varepsilon) < \min\{\log n \frac{K}{4M^2}, k_n I_{C_{11}^2}(M^2)\}. \tag{63}$$

Both terms are increasing in n, as long as $I_{C_{11}^2}(M^2) > 0$. This is true for the C_{11} we consider: if C_{11} is symmetric around zero such that $|C_{11}| < M < \infty$ almost surely, then $I_{C_{11}^2}(M^2) = \infty$, and if C_{mi} is standard normal then $I_{C_{11}^2}(M^2) > 0$.

Therefore, (63) is true when n is large enough, when $I_\infty(\alpha \pm \varepsilon) < \infty$. On the other hand, when $I_\infty(\alpha \pm \varepsilon) = \infty$, then we obtain that the exponential rates converge to infinity, as stated in (40) and (42). We conclude that, for every $\varepsilon > 0$,

$$\liminf_{n\to\infty} -\frac{1}{n}\log \mathbb{P}_{k_n}(\lambda_{\min} \leq \alpha) \geq I_\infty(\alpha + \varepsilon), \tag{64}$$

and

$$\liminf_{n\to\infty} -\frac{1}{n}\log \mathbb{P}_{k_n}(\lambda_{\max} \geq \alpha) \geq I_\infty(\alpha - \varepsilon), \tag{65}$$

and letting $\varepsilon \downarrow 0$ completes the proof for every sequence k_n such that $k_n = o(\frac{n}{\log\log n})$. The proof for $\lambda_{\max}$ is identical to the above proof. □

We believe that the above argument can be extended somewhat further, by making a further split into $K'\log\log n \leq \lambda_{\max} \leq K\log n$ and $\lambda_{\max} \leq K'\log\log n$, but we refrain from writing this down.

3.3 The limiting rate for k large

In this section, we investigate what happens when we take k large. In certain cases, we can show that the rate function, which depends on k, converges to the rate function for Wishart matrices. This will be formulated in the following theorem.

Theorem 3.3 *Assume that C_{ij} satisfies (1), and, moreover, that $\phi_C(t) \leq e^{t^2/2}$ for all t. Then, for all $\alpha \geq 1$, and all $k \geq 2$,*

$$I_k(\alpha) \geq \frac{1}{2}(\alpha - 1 - \log \alpha), \tag{66}$$

and, for all $\alpha \geq 1$,

$$\lim_{k\to\infty} I_k(\alpha) = I_\infty(\alpha) = \frac{1}{2}(\alpha - 1 - \log \alpha). \tag{67}$$

Finally, for all $k_n \to \infty$ such that $k_n = o(\frac{n}{\log\log n})$ and $\alpha \geq 1$,

$$\lim_{n\to\infty} -\frac{1}{n} \log \mathbb{P}_{k_n}(\lambda_{\max} \geq \alpha) = \frac{1}{2}(\alpha - 1 - \log \alpha). \tag{68}$$

Note that, in particular, Theorem 3.3 implies that $I_\infty(\alpha) = \frac{1}{2}(\alpha - 1 - \log \alpha) > 0$ for all $\alpha > 1$.

Theorem 3.3 is a kind of universality result, and shows that, for k large, the rate functions of certain sample covariance matrices converges to the rate function for Wishart matrices. An example where $\phi_C(t) \leq e^{t^2/2}$ holds is when $C_{ij} = \pm 1$ with equal probability. We will call this example the Bernoulli case. A second example is uniform random variable on $[-\sqrt{3}, \sqrt{3}]$, for which also the variance equals 1. We will prove these bounds below.

Of course, the relation that $\phi_C(t) \leq e^{t^2/2}$ for random variables with mean 0 and variance 1, is equivalent to the statement that $\phi_C(t) \leq e^{t^2\sigma^2/2}$ for a random variable C with mean 0 and variance σ^2. Thus, we will check the condition for uniform random variables on $[-1, 1]$ and for the Bernoulli case. We will denote the moment generating functions by ϕ_U and ϕ_B. We start with the second, for which we have that

$$\phi_B(t) = \cosh(t) = \sum_{n=0}^{\infty} \frac{t^{2n}}{(2n)!} \leq \sum_{n=0}^{\infty} \frac{t^{2n}}{2^n n!} = e^{t^2/2}, \tag{69}$$

since $(2n)! \geq 2^n n!$ for all $n \geq 0$. The proof for ϕ_U is similar. Indeed,

$$\phi_U(t) = \frac{\sinh(t)}{t} = \sum_{n=0}^{\infty} \frac{t^{2n}}{(2n+1)!} \leq \sum_{n=0}^{\infty} \frac{t^{2n}}{6^n n!} = e^{t^2/6} = e^{t^2\sigma^2/2}, \tag{70}$$

since now $(2n+1)! \geq 6^n n!$ for all $n \geq 0$.

Proof of Theorem 3.3. Using Theorem 2.1 and 3.1, we claim that it suffices to prove that, uniformly for $\mathbf{x}$ with $\|\mathbf{x}\|_2 = 1$ and $t < 1$,

$$\mathbb{E}[e^{tS^2_{\mathbf{x},i}}] \leq \frac{1}{\sqrt{1-2t}}. \tag{71}$$

We will prove (71) below, and first prove Theorem 3.3 assuming that (71) holds.

When (71) holds, then, using Theorem 3.1 and (5), it immediately follows that (66) holds. Here we also use that $\alpha \mapsto \frac{1}{2}(\alpha - 1 - \log \alpha)$ is continuous, so that the limit over $\varepsilon \downarrow 0$ can be computed.

To prove (67), we take $\mathbf{x} = \frac{1}{\sqrt{k}}(1, \ldots, 1)$, to obtain that, with $S_k = \sum_{i=1}^k C_{i1}$,

$$I_k(\alpha) \leq \sup_t \left(t\alpha - \log \mathbb{E}[e^{\frac{t}{k}S_k^2}] \right). \tag{72}$$

We claim that, when $k \to \infty$, for all $0 \leq t < 1$,

$$\mathbb{E}[e^{\frac{t}{k}S_k^2}] \to \mathbb{E}[e^{tZ^2}] = \frac{1}{\sqrt{1-2t}}, \tag{73}$$

where Z is a standard normal random variable. This implies the lower bound for $I_k(\alpha)$, and thus (67). Equation (68) follows in a similar way, also using that $\alpha \mapsto \frac{1}{2}(\alpha - 1 - \log \alpha)$ is continuous.

We complete the proof by showing that (71) and (73) hold. We start with (71). We rewrite, for $t \geq 0$, and writing Z for a standard normal random variable,

$$\mathbb{E}[e^{tS_{\mathbf{x},i}^2}] = \mathbb{E}[e^{\sqrt{2t}ZS_{\mathbf{x},i}}] = \mathbb{E}\big[\prod_{j=1}^{k} \phi_C(\sqrt{2t}Z\mathbf{x}_j)\big]. \tag{74}$$

We now use that $\phi_C(t) \leq e^{t^2/2}$ to arrive at

$$\mathbb{E}[e^{tS_{\mathbf{x},i}^2}] \leq \mathbb{E}\big[\prod_{j=1}^{k} e^{tZ^2\mathbf{x}_j^2}\big] = \mathbb{E}[e^{tZ^2}] = \frac{1}{\sqrt{1-2t}}. \tag{75}$$

This completes the proof of (71). We proceed with (73). We use

$$\mathbb{E}[e^{tS_k^2}] = \mathbb{E}\big[\prod_{j=1}^{k} \phi_C(\sqrt{\frac{2t}{k}}Z)\big]. \tag{76}$$

We will use dominated convergence. By the assumption, we have that $\phi_C(\sqrt{\frac{2t}{k}}Z) \leq e^{\frac{t}{k}Z^2}$, so that $\prod_{j=1}^{k} \phi_C(\sqrt{\frac{2t}{k}}Z) \leq e^{tZ^2}$, which has a finite expectation when $t < 1/2$. Moreover, $\prod_{j=1}^{k} \phi_C(\sqrt{\frac{2t}{k}}z)$ converges to e^{tz^2} pointwise in z. Therefore, dominated convergence proves the claim in (73), and completes the proof. □

4 The smallest eigenvalue for $C_{ij} = \pm 1$

Unfortunately, we are not able to prove a similar result as in Theorem 3.3 for the smallest eigenvalue. In fact, as we will comment on in more detail in Section 4.2 below, we expect the result to be *false* for the smallest eigenvalue, in particular when α is small. There is one example where we can prove a partial convergence result, and that is when $C_{ij} = \pm 1$ with equal probability. Indeed, in this case it is shown in [15, Section IV] that (71) holds for all $t \geq -1$. This leads to the following result, which also implies that $I_k(\alpha) > 0$ for $\alpha \neq 1$ in the case where $\mathrm{Var}(C_{11}^2) = 0$ (recall also Theorem 2.1):

Theorem 4.1 *Assume that $C_{ij} = \pm 1$ with equal probability. Then, for all $\alpha \geq 1/2$, and all $k \geq 2$,*

$$I_k(\alpha) \geq \frac{1}{2}(\alpha - 1 - \log\alpha), \tag{77}$$

and, for all $\alpha \geq 1/2$,

$$I_\infty(\alpha) = \frac{1}{2}(\alpha - 1 - \log\alpha). \tag{78}$$

Finally, for all $0 < \alpha \leq 1/2$,

$$I_k(\alpha) \geq \frac{1}{2}(-\alpha + \log 2). \tag{79}$$

Proof. The proof of (77–78) is identical to the proof of Theorem 3.3, now using that (71) holds for all $t \geq -1$. Equation (79) follows since $I_k(\alpha) \geq \inf_{\mathbf{x}:\|\mathbf{x}\|_2=1}\Big(-\frac{\alpha}{2} - \log\mathbb{E}[e^{-\frac{1}{2}S_{\mathbf{x},i}^2}]\Big)$ and the bound on the moment generating function for $t = -1$. □

4.1 Rate for the probability of one or more zero eigenvalues for $C_{ij} = \pm 1$

In the above computations, we obtain no control over the probability of a large deviation of the smallest eigenvalue $\lambda_{\min}$. In this and the next section, we investigate this problem in the case where $C_{ij} = \pm 1$.

Proposition 4.2 *Suppose that $C_{ij} = \pm 1$ with equal probability. Then, for all $0 < l \leq k-1$, and any $k_n = \mathcal{O}(n^b)$ for some b,*

$$\lim_{n\to\infty} -\frac{1}{n}\log \mathbb{P}_{k_n}(\lambda_1 = \ldots = \lambda_l = 0) = l\log 2, \tag{80}$$

where the λ_i denote eigenvalues of W arranged in increasing order.

Proof. The upper bound in (80) is simple, since, to have l eigenvalues equal to zero, we can take the first $l+1$ columns of C to be equal. For eigenvectors $\mathbf{w}$ of W,

$$\langle \mathbf{w}, W\mathbf{w}\rangle = \frac{1}{n}\langle \mathbf{w}, CC^T\mathbf{w}\rangle = \frac{1}{n}\|C^T\mathbf{w}\|_2^2,$$

we obtain that $\mathbf{w}$ is an eigenvector with eigenvalue zero precisely when $\|C^T\mathbf{w}\|_2 = 0$. When the first $l+1$ columns of C are equal, then there are l linearly independent vectors for which $\|C^T\mathbf{w}\|_2 = 0$, so that the multiplicity of the eigenvalue zero is at least l. Moreover, the probability that the first $l+1$ columns of C are equal is equal to 2^{nl}.

We prove the lower bound in (80) by induction by showing that

$$\lim_{n\to\infty} -\frac{1}{n}\log \mathbb{P}_{k_n}(\lambda_1 = \ldots = \lambda_l = 0) \geq l\log 2. \tag{81}$$

When $l = 0$, then the claim is trivial. It suffices to advance the induction hypothesis.

Suppose that there are l linear independent eigenvectors with eigenvalue zero. Since the eigenvectors can be chosen to be orthogonal, it is possible to make linear combinations, such that the first $l-1$ all have one zero coordinate j, whereas the l^{th} has all coordinates zero except coordinate j. This means that the first $l-1$ eigenvectors fix some part of C^T, but not the j^{th} column. The l^{th} eigenvector however fixes precisely this column. Fixing one column of C^T has probability 2^{-n}. The number of possible rows j is bounded by k, which is turn is bounded by $n^b = e^{o(n)}$. Therefore, we have

$$\lim_{n\to\infty} -\frac{1}{n}\log \mathbb{P}_{k_n}(\lambda_1 = \ldots = \lambda_l = 0) \geq \log 2 + \lim_{n\to\infty} -\frac{1}{n}\log \mathbb{P}_{k_n}(\lambda_1 = \ldots = \lambda_{l-1} = 0). \tag{82}$$

The claim follows from the induction hypothesis. □

Note that Proposition 4.2 shows that (78) cannot be extended to $\alpha = 0$. Therefore, a changeover takes place between $\alpha = 0$ and $\alpha \geq \frac{1}{2}$, where for $\alpha \geq \frac{1}{2}$, the rate function equals the one for Wishart matrices, while for $\alpha = 0$, this is not the case. We will comment more on this in Conjecture 4.3 below.

4.2 A conjecture about the smallest eigenvalue for $C_{ij} = \pm 1$

We have already shown that (77) is sharp. By Proposition 4.2, (79) is not sharp, since $I_{\min}(0) = \log 2$, whereas (79) only yields $\lim_{\alpha\downarrow 0} I_{\min}(\alpha) \geq \frac{1}{2}\log 2$.

We can use (18) again with $\mathbf{x}' = \frac{1}{\sqrt{2}}(1,1,0,\cdots)$. For this vector, $\mathbb{E}[e^{tS^2_{\mathbf{x}',i}}] = \frac{1}{2}(e^{2t}+1)$, and calculating the according rate function gives $I_k(\alpha) \leq I^{(2)}(\alpha) = \frac{\alpha}{2}\log\alpha + \frac{2-\alpha}{2}\log(2-\alpha)$, which implies that $\lim_{\alpha\downarrow 0} I_k(\alpha) \leq \log 2$.

It appears that below a certain $\alpha = \alpha_k^*$, the optimal strategy changes from $\mathbf{x}^{(k)} = \frac{1}{\sqrt{k}}(1,1,\cdots)$ to $\mathbf{x}^{(2)} = \frac{1}{\sqrt{2}}(1,1,0,\cdots)$. In words, that means that for not too small eigenvalues of W, all entries of C contribute equally to create a small eigenvalue. However, smaller values of the smallest eigenvalues of W are created by only two columns of C, whereas the others are close to orthogonal. Thus, a change in strategy occurs, which gives rise to a phase transition in the asymptotic exponential rate for the smallest eigenvalue.

We have the following conjecture:

Conjecture 4.3 *For each k and all $\alpha \geq 0$, there exists an $\alpha^* = \alpha_k^* > 0$ so that*

$$I_k(\alpha) = I^{(2)}(\alpha) = \frac{\alpha}{2}\log\alpha + \frac{(2-\alpha)}{2}\log(2-\alpha),$$

for $\alpha \leq \alpha_k^*$*. For* $\alpha > \alpha_k^*$*,*

$$I^{(2)}(\alpha) > I_k(\alpha) \geq I_\infty(\alpha) = \frac{1}{2}(\alpha - 1 - \log \alpha).$$

For $k \to \infty$*, the last inequality will become an equality. Consequently,* $\lim_{k\to\infty} \alpha_k^* = \alpha^*$*, which is the positive solution of* $I_\infty(\alpha) = I^{(2)}(\alpha)$*.*

For $k = 2$, the conjecture is trivially true, since the two optimal strategies are the same, and the only possible. Note that in the proof of Proposition 4.2, we used that to have a zero eigenvalue, we need two vectors of C to be equal. Thus, the conjecture is also proven for $\alpha = 0$. Furthermore, with Theorem 3.2 and (46), the conjecture follows for all k for $\alpha \geq \frac{1}{2}$. We lack a proof for $0 < \alpha < \frac{1}{2}$. Numerical evaluation gives that $\alpha_3^* \approx 0.425$, and for $k \to \infty$, $\alpha_k^* \approx 0.253$. We have some evidence that suggests that α_k^* decreases with k.

5 An application: Mobile Communication Systems

Our results on the eigenvalues of sample covariance matrices was triggered by a problem in mobile communication systems. In this case, we take the C matrix as a coding sequence, for which we can assume that the elements are ± 1. Thus, all our results apply to this case. In this section, we will describe the consequences of our results on this problem.

5.1 Soft-Decision Parallel Interference Cancellation

In Code Division Multiple Access (CDMA) communication systems, each of k users multiplies his data signal by an individual coding sequence. The base station can distinguish the different messages by taking the inner product of the total signal with each coding sequence. This is called Matched Filter (MF) decoding. An important application is mobile telephony. Since, due to synchronisation problems, it is unfeasible to implement completely orthogonal codes for mobile users, the decoded messages will suffer from Multiple Access Interference (MAI). In practice, pseudo-random codes are used. Designers of decoding schemes are interested in the probability that a decoding error is made.

In the following, we explain a specific method to iteratively estimate and subtract the MAI, namely, Soft Decision Parallel Interference Cancellation (SD-PIC). For more background on SD-PIC, see [5], [8], [10] and [18], as well as the references therein. Because this procedure is linear, it can be expressed in matrix notation. We will show that the possibility of a decoding error is related to a large deviation of the maximum or minimum eigenvalue of the code correlation matrix.

To highlight the aspects that are relevant in this article, we suppose that each user sends only one data bit $b_m \in \{+1, -1\}$, and we omit noise from additional sources. We can denote all sent data multiplied by their amplitude in a column vector $\mathbf{Z}$, i.e., $\mathbf{Z}_m = \sqrt{P_m} b_m$, where P_m is the power of the m^{th} user. The k codes are modeled as the different rows of length n of the code matrix C, consisting of i.i.d. random bits with distribution

$$\mathbb{P}(C_{mi} = +1) = \mathbb{P}(C_{mi} = -1) = \frac{1}{2}.$$

Thus, k plays the role of the number of users, while n is the length of the different codes.

The base station then receives a total signal $\mathbf{s} = C^T \mathbf{Z}$. Decoding for user m is done by taking the inner product with the code of the m^{th} user $(C_{1m}, \ldots, C_{nm})$, and dividing by n. This yields an estimate $\hat{\mathbf{Z}}_m^{(1)}$ for the sent signal $\mathbf{Z}_m$. In matrix notation, the vector $\mathbf{Z}$ is estimated by

$$\hat{\mathbf{Z}}^{(1)} = \frac{1}{n} C \mathbf{s} = W \mathbf{Z}.$$

Thus, we see that multiplying with the matrix W is equivalent to the MF decoding scheme. In order to estimate the signal, we must find the inverse matrix W^{-1}. From $\hat{\mathbf{Z}}^{(1)}$, we estimate the sent bit b_m by

$$\hat{b}_m^{(1)} = \text{sign}(\hat{\mathbf{Z}}_m^{(1)}) \tag{83}$$

(where, when $\hat{\mathbf{Z}}_m^{(1)} = 0$, we toss an independent fair coin to decide what the value of $\text{sign}(\hat{\mathbf{Z}}_m^{(1)})$ is). Below, we explain the role of the eigenvalues of W in the more advanced SD-PIC decoding scheme.

The MF estimate contains MAI. When we write $\hat{\mathbf{Z}} = \mathbf{Z} + (W - I)\mathbf{Z}$, it is clear that the estimated bit vector is a sum of the correct bit vector and MAI. In SD-PIC, the second term is subtracted, with $\mathbf{Z}$ replaced by $\hat{\mathbf{Z}}$. In the case of multistage PIC, each new estimate is used in the next PIC iteration. We will now write the multistage SD-PIC procedure in matrix notation. We number the successive SD estimates for $\mathbf{Z}$ with an index s, where $s = 1$ corresponds to the MF decoding. In each new iteration, the latest guess for the MAI is subtracted. The iteration in a recursive form is therefore:

$$\hat{\mathbf{Z}}^{(s)} = \hat{\mathbf{Z}}^{(1)} - (W - I)\hat{\mathbf{Z}}^{(s-1)}. \tag{84}$$

This can be worked out to

$$\hat{\mathbf{Z}}^{(s)} = \sum_{\varsigma=0}^{s-1}(I - W)^{\varsigma} W\mathbf{Z}. \tag{85}$$

We then estimate b_m by

$$\hat{b}_m^{(s)} = \mathrm{sign}(\hat{\mathbf{Z}}_m^{(s)}). \tag{86}$$

When $s \to \infty$, the series $\sum_{\varsigma=0}^{s-1}(I - W)^{\varsigma}$ converges to W^{-1}, as long as the eigenvalues of W are between 0 and 2. Otherwise, a decoding error is made. This is the crux to the method, see also [21] for the above matrix computations. When $k = o(n/\log\log n)$, the values $\lambda_{\min} = 0$ and $\lambda_{\max} \geq 2$ are large deviations, and therefore our derived rate functions provide information on the error probability. In the next section, we will describe these results, and we will also obtain bounds on the exponential rate of a bit error in the case that s is fixed and k is large. For an extensive introduction to CDMA and PIC procedures, we refer to [18].

5.2 Results for Soft-Decision Parallel Interference Cancelation

There are two cases that need to be distinguished, namely, the case where $s \to \infty$, and the case where s is fixed. We start with the former, which is simplest. As explained in the previous section, due to the absence of noise, there can only be bit-errors when $\lambda_{\min} = 0$ or when $\lambda_{\max} \geq 2$. By (79), the rate of $\lambda_{\min} = 0$ is at least $\frac{1}{2}\log 2 \approx 0.35...$, whereas the rate of $\lambda_{\max} \geq 2$ is bounded below by $\frac{1}{2} - \frac{1}{2}\log 2 \approx 0.15...$ The latter bound is weaker, and thus, by the largest-exponent-wins principle, we obtain the following result:

Theorem 5.1 (Bit-error rate for optimal SD-PIC) *For all k fixed, or for $k = k_n \to \infty$ such that $k_n = o(\frac{n}{\log\log n})$,*

$$-\frac{1}{n}\log \mathbb{P}_k\big(\exists m = 1, \ldots, k \textit{ for which } \lim_{s\to\infty} \hat{b}_m^{(s)} \neq b_m\big) \geq \frac{1}{2} - \frac{1}{2}\log 2. \tag{87}$$

We emphasize that in the statement of the result, we write that $\lim_{s\to\infty} \hat{b}_m^{(s)} \neq b_m \forall m = 1, \ldots, k$ for the statement that either $\lim_{s\to\infty} \hat{b}_m^{(s)}$ does not exist, or that $\lim_{s\to\infty} \hat{b}_m^{(s)}$ exists, but is unequal to b_m. We observe that when $\lambda_{\max} > 2$, then

$$\hat{\mathbf{Z}}^{(s)} = \sum_{\varsigma=0}^{s-1}(I - W)^{\varsigma} W\mathbf{Z} \tag{88}$$

oscillates, so that we can expect there to be errors in every stage. This is sometimes called the *ping-pong effect* (see [21]). Thus, one would expect that

$$-\frac{1}{n}\log \mathbb{P}_{k_n}\big(\exists m = 1, \ldots, k \text{ for which } \lim_{s\to\infty} \hat{b}_m^{(s)} \neq b_m\big) = \frac{1}{2} - \frac{1}{2}\log 2.$$

However, this depends also on the relation between $\mathbf{Z}$ and the eigenvector corresponding to $\lambda_{\max}$. Indeed, when $\mathbf{Z}$ is orthogonal to the eigenvector corresponding to $\lambda_{\max}$, then the equality does not follow. To avoid such problems, we stick to lower bounds on the rates in this section, rather than asymptotics.

We next go to the case where s is fixed. We again consider the case where k is large and fixed, or that $k = k_n \to \infty$. In this case, it can be expected that the rate converges to 0 as $k \to \infty$. We already know that the probability that $\lambda_{\max} \geq 2$ or $\lambda_{\min} = 0$ is exponentially small with *fixed* strictly

positive lower bound on the exponential rate. Thus, we shall assume that $0 < \lambda_{\min} \leq \lambda_{\max} < 2$. We can then rewrite

$$\hat{\mathbf{Z}}^{(s)} = \sum_{\varsigma=0}^{s-1} (I-W)^{\varsigma} W \mathbf{Z} = \big[I - (I-W)^s\big]\mathbf{Z}. \tag{89}$$

For simplicity, we will first assume that $\mathbf{Z}_i = \pm 1$ for all $i = 1, \ldots, k$, which is equivalent to assuming that all powers are equal. When s is fixed, we cannot have any bit-errors when

$$\Big|\big((I-W)^s \mathbf{Z}\big)_i\Big| < 1. \tag{90}$$

We can bound

$$\Big|\big((I-W)^s \mathbf{Z}\big)_i\Big| \leq \varepsilon_k^s \|\mathbf{b}\|_2, \tag{91}$$

where $\varepsilon_k = \max\{1-\lambda_{\min}, \lambda_{\max}-1\}$. Since $\|\mathbf{b}\|_2 = \sqrt{k}$, we obtain that there cannot be any bit-errors when $\varepsilon_k^s \sqrt{k} < 1$. This gives an explicit relation between the bit-errors and the eigenvalues of a random sample covariance matrix. By applying the results from the previous two sections, we obtain the following theorem:

Theorem 5.2 (Bit-error rate for finite-stage SD-PIC and k fixed) *For all k such that $k > 2^{2s}$,*

$$-\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_k\big(\exists m = 1, \ldots, k \textit{ for which } \lim_{s\to\infty} \hat{b}_m^{(s)} \neq b_m\big) \geq \frac{1}{4\sqrt[s]{k}}\Big(1 + \mathcal{O}\big(\frac{1}{\sqrt[s]{k}}\big)\Big). \tag{92}$$

When the signals are different, related results can be obtain in terms of the minimal and maximal element of $\mathbf{Z}$. We will not write this case out.

Proof. By the computation in (91), there can be no bit-errors when $1-\lambda_{\min}$ and $\lambda_{\max}-1$ are both at most $1/\sqrt[2s]{k}$. Thus,

$$\mathbb{P}_k\big(\exists m = 1, \ldots, k \text{ for which } \lim_{s\to\infty} \hat{b}_m^{(s)} \neq b_m\big) \leq \mathbb{P}_k\big(\lambda_{\min} \leq 1 - \frac{1}{\sqrt[2s]{k}}\big) + \mathbb{P}_k\big(\lambda_{\max} \geq 1 + \frac{1}{\sqrt[2s]{k}}\big). \tag{93}$$

Each of these terms is bounded by, using Theorem 2.1,

$$e^{-n \min\Big\{\lim_{\varepsilon\downarrow 0} I_k\big(1 - \frac{2}{2\sqrt[s]{k}} - \varepsilon\big), I_k\big(1 + \frac{2}{2\sqrt[s]{k}} + \varepsilon\big)\Big\}(1+o(1))}. \tag{94}$$

Since, by Theorem 4.1 and $\alpha \geq \frac{1}{2}$, we have that $I_k(\alpha) \geq I_\infty(\alpha) = \frac{1}{2}(\alpha - 1 - \log\alpha)$, and

$$I_\infty(\alpha) = \frac{1}{4}(\alpha-1)^2 + \mathcal{O}(|\alpha-1|^3), \tag{95}$$

the result follows when k is so large that $1 - \frac{1}{\sqrt[2s]{k}} > \frac{1}{2}$. The latter is equivalent to $k > 2^{2s}$. □

We finally state a result that applied to $k = k_n$:

Theorem 5.3 (Bit-error rate for finite-stage SD-PIC and $k = k_n$) *For $k_n = o\big(\frac{n^{\frac{s}{s+1}}}{\log n}\big)$,*

$$-\frac{\sqrt[s]{k_n}}{n} \log \mathbb{P}_{k_n}\big(\exists m = 1, \ldots, k \textit{ for which } \lim_{s\to\infty} \hat{b}_m^{(s)} \neq b_m\big) \geq \frac{1}{4} + \mathcal{O}\big(\frac{1}{\sqrt[s]{k_n}}\big). \tag{96}$$

Proof. We use (93), to conclude that we need to derive bounds for $\mathbb{P}_{k_n}\big(\lambda_{\min} \leq 1 - \frac{1}{\sqrt[2s]{k_n}}\big)$ and $\mathbb{P}_{k_n}\big(\lambda_{\max} \geq 1 + \frac{1}{\sqrt[2s]{k_n}}\big)$. Unfortunately, the bounds $1 - \frac{1}{\sqrt[2s]{k_n}}$ and $1 + \frac{1}{\sqrt[2s]{k_n}}$ on the smallest and largest eigenvalues depend on n, rather than being fixed. Therefore, we need to adapt the proof of Theorem 3.1.

We note that, by Theorem 3.1,

$$\mathbb{P}_{k_n}(\lambda_{\max} \geq 2) = e^{-(\frac{1}{2} - \frac{1}{2}\log 2)n(1+o(1))}.$$

Then, we use (60) with $\Omega_n = 2$, and choose R_n such that

$$\frac{2}{R_n} = o\Big(\frac{1}{\sqrt[2s]{k_n}}\Big), \tag{97}$$

so that $R_n \gg \sqrt[2s]{k_n}$. Applying (59) and (54), we see that we need that $R_n^{k_n} = e^{o\big(\frac{n}{\sqrt[s]{k_n}}\big)}$, so that $k_n = o\big(\frac{n^{\frac{s}{s+1}}}{\log n}\big)$ is sufficient. Finally, by Theorem 4.1 and (95), we have that

$$I_{k_n}\big(1 \pm \frac{1}{\sqrt[2s]{k_n}}\big) \geq \frac{1}{4\sqrt[s]{k_n}}\Big(1+\mathcal{O}\big(\frac{1}{\sqrt[s]{k_n}}\big)\Big). \tag{98}$$

This completes the proof. □

We now discuss the above results. In [14], it was conjectured that when $s = 2$, the rate of a *single* bit error for a fixed user is asymptotic to $\frac{1}{2\sqrt{k}}$ when $k \to \infty$. See also [13]. We see that we obtain a similar result, but our constant is $1/4$ rather than the expected $1/2$. On the other hand, our result is valid for *all* $s \geq 2$.

Related results where obtained for a related model, *Hard-Decision Parallel Interference Cancelation* (HD-PIC) where bits are iteratively estimated by bits, i.e., the estimates are rounded to ± 1. Thus, this scheme is not linear, as SD-PIC is. In [15, 16], similar results as the above are obtained, and it is shown that the rate for a bit-error for a given user is asymptotic to $\frac{s}{8}\sqrt[s]{\frac{4}{k}}$ when s is fixed and $k \to \infty$. This result is similar in spirit as the one in Theorem 5.2 above. The explanation of why the rate tends to zero as $1/\sqrt[s]{k}$ is much simpler for the case of SD-PIC, where the relation to eigenvalues is rather direct, compared to the explanation for HD-PIC, which is much more elaborate. It is interesting to see that both when $s = \infty$ and when s is finite and $k \to \infty$, the rates in the two systems are of the same order.

Interestingly, in [17], it was shown that for $s = 1$ and $k_n = \frac{n}{\gamma \log n}$, with high probability, all bits are estimated correctly when $\gamma < 2$, while, with high probability, there is at least one bit-error when $\gamma > 2$. Thus, $k_n = \mathcal{O}(\frac{n}{\log n})$ is critical for the MF system, where we do not apply SD-PIC. For SD-PIC with an arbitrary number of staged of SD-PIC, we have no bit-errors with large probability for all $k_n = \frac{n}{\gamma \log n}$ for *all* $\gamma > 0$, and we can even pick larger values of k_n such that $k_n = o(\frac{n}{\log\log n})$. Thus, SD-PIC is more efficient than MF, in the sense that it allows more users to transmit without creating bit-errors. Furthermore, in [17], the results proved in this paper are used for a further comparison between SD-PIC, HD-PIC and MF. Unfortunately, when we only apply a *finite* number of stages of SD-PIC, we can only allow for $k_n = o\big(\frac{n^{\frac{s}{s+1}}}{\log n}\big)$ users. Similar results were obtained for HD-PIC when $k_n = o(\frac{n}{\log n})$.

We close this discussion on SD-PIC and HD-PIC by noting that for $k = \beta n$, $\lambda_{\min}$ converges to $(1-\sqrt{\beta})_+^2$, while the largest eigenvalue $\lambda_{\max}$ converges to $(1+\sqrt{\beta})^2$ (see [3, 4, 23]). This is explained in more detail in [8], and illustrates that SD-PIC has no bit-errors with probability converging to 1 whenever $\beta < (\sqrt{2}-1)^2 \approx 0.17...$ However, unlike the case where $k_n = o(\frac{n}{\log\log n})$, we do not obtain bounds on how the probability of a bit-error tends to zero.

A further CDMA system is the *decorrelator*, which explicitly inverts the matrix W (without approximating it by the partial sum $\sum_{\varsigma=0}^{s-1}(I-W)^{\varsigma}$). One way of doing so is to fix a large value M and to compute

$$\hat{\mathbf{Z}}_M^{(s)} = M^{-1}\sum_{\varsigma=0}^{s-1}\big(I - \frac{W}{M}\big)^{\varsigma} W\mathbf{Z}, \tag{99}$$

and

$$\hat{b}_{m,M}^{(s)} = \mathrm{sign}(\hat{\mathbf{Z}}_{m,M}^{(s)}). \tag{100}$$

This is a certain *weighted* SD-PIC scheme. This scheme will converge to $\mathbf{b}$ as $s \to \infty$ whenever $\lambda_{\min} > 0$ and $\lambda_{\max} < M$. By taking M such that $I_\infty(M) \geq \log 2$, and using Proposition 4.2, we obtain the following result:

Theorem 5.4 (Bit-error rate for optimal weighted SD-PIC) *For all k fixed, or for $k = k_n \to \infty$ such that $k_n = o(\frac{n}{\log\log n})$ and M such that $I_\infty(M) \geq \log 2$,*

$$-\frac{1}{n}\log \mathbb{P}_k\big(\exists m = 1, \ldots, k \text{ for which } \hat{b}_{m,M}^{(s)} \neq b_m\big) \geq \log 2. \tag{101}$$

The above result can even be generalised to k_n that grow arbitrarily fast with n, by taking M dependent on n. For example, when we take $M > k_n$, then $\lambda_{\max} \leq k_n < M$ is guaranteed.

Further interesting problems arise when we allow the received signal to be *noisy*. In this case, the bit-error can be caused either by the properties of the eigenvalues, as in the case when there is no noise, or by the noise. When there is noise, weighted SD-PIC for large M *enhances* the noise, which makes the problem significantly harder. See [18] for further details. A solution to Conjecture 4.3 may prove to be useful in such an analysis.

Acknowledgement. The work of AF and RvdH was supported in part by NWO. The work of RvdH and MJK was performed in part at Delft University of Technology.

References

[1] Z. D. Bai and J. W. Silverstein (1998). No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices, *Ann. Probab.*, **26**, 316–345.

[2] Z. D. Bai and J. W. Silverstein (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices, *Ann. Probab.*, **32**, 553–605.

[3] Z.D. Bai, J.W. Silverstein and Y.Q. Yin (1988). A note on the largest eigenvalue of a large dimensional sample covariance matrix, *J. Multivariate Anal.*, **26**, 166-168.

[4] Z.D. Bai and Y.Q. Yin (1993). Limit of the smallest eigenvalue of a large dimensional sample covariance matrix, *Ann. Probab.*, **21**, 1275-1294.

[5] R.M. Buehrer, S.P. Nicoloso, and S.Gollamudi (June 1999). Linear versus Non-linear Interference Cancellation, *J. Commun. and Networks*, **1**, 2, 118-133.

[6] A. Dembo and O. Zeitouni (1998). Large Deviations Techniques and Applications, *Springer-Verlag, New York.*

[7] A.C. Fey-den Boer, R. van der Hofstad, M.J. Klok (2003). Linear Interference Cancellation in CDMA Systems and Large Deviations of the Correlation Matrix Eigenvalues, *Proc. Symp. IEEE Benelux Chapter on Comm. and Veh. Techn. 2003, Eindhoven.*

[8] A. Grant, and C. Schlegel (2001). Convergence of Linear Interference Cancellation Multiuser Receivers, *IEEE Trans. on Comm.*, **49**, 10, 1824-1834.

[9] A. Guionnet (2002). Large deviation asymptotics for spherical integrals, *Journ. of Funct. Anal.*, **188**, 461-515.

[10] D. Guo, L.K. Rasmussen, S. Tun, and T.J. Lim (Jan. 2000). A Matrix-Algebraic Approach to Linear Parallel Interference Cancellation in CDMA, *IEEE Trans. Comm.*, **48**, 1, 152-161.

[11] F. Hiai, D. Petz (1998). Eigenvalue density of the Wishart matrix and large deviations, *Inf. Dim. Anal. Quantum Prob.*, **1**, 633-646.

[12] F. den Hollander (2000). Large Deviations, *Fields Monograph Series, AMS, USA.*

[13] R. van der Hofstad, G. Hooghiemstra and M.J. Klok (2002a). Large deviations for code division multiple access systems, SIAM Journal of Applied Mathematics **62**, no. 3: 1044–1065.

[14] R. van der Hofstad, G. Hooghiemstra and M.J. Klok (2002b). Analytical methods for CDMA systems with parallel interference cancellation: the large deviation approach, Wireless Personal Communication, **21**: 289-307.

[15] R. van der Hofstad and M.J. Klok (2003). Performance of DS-CDMA Systems with Optimal Hard Decision Parallel Interference Cancellation, *IEEE Trans. on Information Theory* **25**, Nr 11: 2918-2940.

[16] R. van der Hofstad and M.J. Klok (2004). Improving the performance of 3G communication systems: the hard decision case, *Adv. Appl. Prob.* **36**: 1046-1084.

[17] R. van der Hofstad, M. Löwe and F. Vermet (2006). The effect of system load on the existence of bit-errors in CDMA with and without parallel interference cancelation, *IEEE Transactions on Information Theory Correspondence* **52**: 4733–4741.

[18] Klok M.J. (2001). *Performance Analysis of Advanced Third Generation Receivers*, Ph.D. thesis Delft University of Technology, the Netherlands.

[19] V.A. Marchenko and L.A. Pastur (1967). The distribution of eigenvalues in certain sets of random matrices, *Mat. Sb.*, **72**, 507-536.

[20] E. Olivieri and M.E. Vares (2005). *Large deviations and metastability*, Encyclopedia of Mathematics and its Applications, Cambridge University Press, Cambridge.

[21] L.K. Rasmussen and I.J. Oppermann (2003). Ping-pong effects in linear parallel interference cancellation for CDMA, IEEE Transactions on Wireless Communications **2** no. 2: 357–363.

[22] C.A. Rogers (1963). Covering a Sphere with Spheres, *Mathematika*, **10**, 157-164.

[23] Y. Q. Yin, Z. D. Bai, and P. R. Krishnaiah (1988). On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix, *Probab. Theory Related Fields*, **78**, 509–521.